代理-自我混同体：一个不可还原的新命名

作者 鲍锦朝 · SDE 学员 · 约 3.0 万字 · 发表于2026年7月24日

摘 要 · ABSTRACT

本文从一道被逼出的裂缝出发，执行一项发生学工序。那个裂缝位于“伪发生”诊断的底层：它将“亲自”与“代偿”预设为在同一个认知行为中互斥的两端，由此切开了“真发生”与“伪发生”的边界。本文对这口深井发动了一次更底层的追问：当这个互斥预设有朝一日在某些人的真实认知生活中不再可操作时，从那道被切开的裂缝中结算出来的，将不再是“真”与“伪”的对立，而是一个此前不被任何已有命名所独占的第三稳定态。本文将其命名为代理-自我混同体。它描述的状态是：代理能力已被内化为自我构成的有机成分，主体无法也无须区分认知产出中“我”与“它”的边界，却由此生发出一种真诚的、不可被旧判别标准戳破的效能体验。本文从发生机制上推演了这一新结构是如何从制度与技术的持续对撞中被结算出来的，以教育学现场为例给出了一场卷入制度-个体实际碰撞的写生、一个微观深描与可操作判准，完成了它与分布式认知、认知外包、自我效能感等邻近概念的不可还原性切割，正面回应了三个最强反驳（含一个深化的自反性审判，以及伪发生作为独立反驳者的正面对质），并在结论中为这一命名给出了明确的、可被短期与长期实证裁决的证伪条件。在承认混同体之后，本文并不给出“解决方案”，而是揭示：教育评价、制度设计、代际公平将面临一轮此前不曾被提出的新冲突，而这场冲突没有终点。

关键词：关键词：代理-自我混同体；伪发生；不可代偿性；效能嵌入；发生学

一、引言：从一道被逼出的裂缝开始

2020年代中期，关于AI经济效应的一个锋利诊断浮出水面。这个诊断指出，AI时代的经济系统中出现了一类全新的交易对象。它不是商品，不是服务，而是“发生了”这个事件的证明。能证明一个学生经历过独立思考的反思报告、能证明一个管理者做出了艰难决策的策略叙事、能证明一个人获得了心灵成长的体验书写——这些证明可以被AI无限地、低成本地、高质量地生产出来，并从中兑现真实的经济价值。而那个被预设为交易自然伴生物的东西——承受不可逆代价、改写自身结构的“真发生”——在这个产业链中，被系统性地跳过了。

那个诊断被命名为“伪发生”。它的锐利在于，它不是道德控诉，而是一份关于结构性病变的发生学判断：当生产“发生证明”比制造真实发生更高效、更符合消费者对效能的渴望时，市场会诚实且理性地将后者淘汰。它以“不可代偿的自我卷入判决”为控制变量，与鲍德里亚的拟像论（Baudrillard， 1981）、阿克洛夫的柠檬市场（Akerlof， 1970）、凡勃伦的炫耀性消费（Veblen，1899）等既有理论资源之间，拉出了不可还原的理论边界。它让我们看见，AI时代的主体性危机，并不仅仅是“人变懒了”或“技术太强了”，而是一整套能够产出“发生了”全部可观测信号、却系统性跳过主体不可逆自我更新的产业链，已经悄悄嵌入了教育、职场、文化消费的毛细血管。

然而，就在这个概念划下边界的那最后一笔动作中，一道新的裂缝露了出来。裂缝不是来自外部批评者，而是从诊断自身的底层预设中生出来的。

这个底层预设是：不可代偿的自我卷入判决，只能由主体亲自从内部发动；任何外部助力的深度介入，都是对这一判决的削弱乃至剥夺。简言之，亲自与代偿，在同一个认知或决策行为中，是互斥的。 没有这一预设，“伪发生”概念就无法在逻辑上将自己与“真发生”区分开来——它必须有一条判别线，而这条线划在了“亲自”和“代偿”之间。

但如果我们不把这条线当终点，而是对它发动一次更底层的追问呢?

试想这样一个人：他已经无法在“我亲自想到的”与“AI帮我想到的”之间做出清晰切割，但他仍然真诚地、不可否认地体验到了强烈的“我在想”的效能感。那么此刻在他身上发生的，究竟是什么?它不符合旧哲学意义上那个不受沾染的、自行发动的“真发生”。但它也绝不是“伪发生”——因为伪发生被定义为一种事后可辨识的“你其实没有真正发生”的空洞，其诊断前提是存在可被第三方辨识的判别标准。而在这个人身上，我们找不到“他其实知道自己在装”的证据。他不是在装。他完全信。他的效能感是真实的、不可被戳破的，因为戳破它所需的那个“你没发生”的证据，根本不存在于任何第三方观测里，甚至也不存在于他自己的内省里。他已经无法反身去区分“我”和“它”在认知产出中的边界了——边界被代理融化了。

那么，在他身上被发生出来的那个东西，到底该叫什么名字?

这个追问，逼出了本文试图完成的一次激进行动：暂不把“伪发生”当作最终的诊断，而是将它当作一个从特定预设中诞生的产物，回追它诞生之前那片尚未被它命名的地带。 在这片地带里，在“亲自”与“代偿”的互斥预设被解开之后，一个新的显露态从矛盾的结算中浮了出来。本文将它命名为“代理-自我混同体”。伪发生那一刀切下去，将世界分成“真发生”与“伪发生”两块；而本文要告诉读者，在那条切线的正下方，存在着一个不曾被任何已有概念独占的第三种形态。它既不在“亲自”那一侧，也不在“被代偿”那一侧。它就在被刀切开的那个切口上，自己长成了一个独立的东西。

这一概念试图命名的，是一种这样的状态：AI的代理能力——它的文本生成、它的逻辑梳理、它的叙事建构——已经不再被使用者体验为“外部从旁协助的服务”，而是被内化为自我认知机能的一部分，就像记忆、直觉、语感一样自然、透明、不可割除。使用者无法说出“哪一步是代偿，哪一步是亲自”，因为代偿已经不经过他的自觉监控，就融入了他的“我想”之中。由此产生的效能感——“我写出了一篇好论文”“我做出了一个深刻判断”“我完成了一次艰难的自我分析”——是真诚的。它是真的。它不是诈欺，不是表演，不是消费者被蒙蔽。它是一个不可逆的、不可以被旧有的真-伪二分法切开的混合体状态。这个状态，正是“伪发生”概念试图照亮、却又被它自己的预设所遮蔽的一个独立的结构事实。

在展开论证之前，有必要先做一笔方法论说明。本文的核心概念——“代理-自我混同体”——是通过纯粹的发生学推演来结算的。这一推演从“亲自与代偿互斥”这一底层预设的内部裂缝出发，追问旧预设失效后可能结算出的新显露态；此推演不依赖任何特定案例来担保其有效性。本文随后出现的晓晴案例以及一场制度与个体碰撞的写生，仅作为“概念示踪”——帮助读者在人类经验的层面上看见

这个概念所指向的实在，而非为概念提供归纳基础。概念的成立独立于案例；案例只是帮读者看见它的光彩。

以“代理-自我混同体”为中心，本文将走完以下推进路径：第二节，完成从撤销先验到新结构结算的完整工序，给出操控定义与识别判准，并明确标注这些判准是混同体的剖面照片而非生成故事。第三节，推演混同体是如何从制度与技术的持续对撞中被结算出来的——这一推演不依赖任何特定案例，而是对旧稳态裂缝→制度补丁与反补丁的军备竞赛→新结构缓慢滋生的发生学链条的纯概念推导；在此推导之后，插入一场卷入制度-个体实际碰撞的写生，来钉死“混同体的诞生不是一个光滑的认知习惯转变，而是一场包含恐慌、防御、自责与外部含混奖罚的悄然收编”；最后呈现一个微观深描（晓晴案例）及其故障时刻，以完成从概念推演到经验可指认的全程示踪。第四节，将混同体与分布式认知、认知外包、自我效能感等邻近概念逐一切割，完成不可还原性验证，特别回应分布式认知光谱上极高耦合端是否与混同体重叠这一最强挑战。第五节，直面三个最强反驳，包括伪发生作为独立反驳者的正面对质（给出其核心论点的直接引述并逼混同体与之碰撞）、深化版自反性审判、以及与伪发生的操作化区分。第六节，以教育评估为主场域展开跨学科推衍，并向前再推一步——揭示新评价体系自身将催生的新一轮矛盾。第七节，给出混同体概念对“我们与世界的关系”的不可逆改写，以及可被短期与长期实证裁决的证伪条件。本文的终点不是“解决方案”，而是揭示承认混同体之后，冲突将如何在新一层展开。

二、源起：一个不可还原的新命名

2.1 从一道裂缝中长出来

任何一种诊断，在它划出“病症”与“健康”的边界时，都必然在某处停住，把那处预设为理所当然的参照点。“伪发生”诊断的参照点，是那个被AI跳过的、不可代偿的自我卷入判决。但这本身还不是最深的那口井。要触及更深的那口井，必须追问：为什么“不可代偿的自我卷入判决”，就一定是可以被独立识别、独立验证、独立与非代偿状态区分开的？这个预设，比“不可代偿”本身埋得更深。

它可以被表述为这样一句话：亲自与代偿，在同一个认知或决策行为中，是互斥的。要么是主体亲自发动了判决，要么是判决被外部代偿了；二者不可能在同一个认知过程中同时为真。你可以先让AI帮你捋一遍思路，然后自己再做判断——那只是两步，各归各的。但你不能在“捋”的那一步里，说自己既捋了又没捋，既是自己又是AI。这条互斥律，是整个“伪发生”诊断得以立足的逻辑地基。

抓住这个在地基以下沉默运转的预设之后，一件要紧的事发生了：一旦这个预设有朝一日不再成立——不是被论证为“错”，而是在某些人的真实认知生活中不再可操作——那么，被诊断为“伪发生”的那整套现象，就不再只能用“真/伪”二分来切割。一种新的可能冒了出来：在代理深度参与认知过程之后，主体可能真诚地、不可逆地进入一种代理-自我边界模糊的混合态，而他所体验到的效能感，既不是对应于某个实际发生的“真”，也不是对应于一个可以戳穿的“伪”——它本身就是一个新的“真”，是一种以混同为发生条件的真实效能体验。

可以设想一个极限情境来钉住这个判断。一个学生完全在AI的辅助下完成了一篇高质量的批判性思维论文。这一过程已被“伪发生”诊断为一次完美的空洞——产出达标，发生缺席。但这个学生不这么认为。他有自己的证据：他确实感到了强烈的认知唤起——他对选题有过迟疑，他纠正过AI的某个表述因为他觉得“不对味”，他在读到自己论文的某一处时被自己的论证说服了。他无法告诉你“这几步里，哪一步是我亲自想的，哪一步是AI替我生成的”。它们在他的记忆里是融合的、连续的、没有接口的。他的效能感不是伪的。它是真的，而且比他在传统课程里写完一篇磕磕绊绊的短文时更强烈。

在“亲自与代偿互斥”的预设下，我们会说：他被蒙蔽了。他的效能感来自AI替他建构的误认。但在撤销这个预设之后，我们看见了别的东西。我们看见，他的代理使用行为，已经不再是一种“请求外援”——外援意味着我站在内，它在外的界限。它变成了他认知活动的一个不被意识标注为“非我”的延伸层，就像我们不区分“我的手在写”和“我的大脑在想”，我们只感到“我在写”。他的手当然是外部的——它可以被截肢。但在认知体验上，它被编织进了“我”的统一场中，不经过任何刻意的“调用”动作就共同运作。那篇论文，就是他的认知场与AI的生成场协同运作的结果。他不区分它们。他不需要区分它们。他也没有能力去区分它们。

在这种状态下，那一套建立在“我本应亲自想”和“它帮我替代了想”之间的判别框架，碰到了一个搅浑了边界的实在。这个实在，既不能被称为“真发生”——因为它不符合旧哲学中那个不受沾染的

主体判决的标准；也不能被称为“伪发生”——因为它并不建筑在“可以被戳破的虚假”之上。它不是欺骗。它是混成。

这里需要警惕一种可能的误读。当我说“旧判别标准失效”时，我不是在做认识论上的让步——不是在说“因为我们测不准，所以我们只好说它是真的”。恰好相反：旧判别标准的失效，本身就是混同体结构成立的本体论证据。旧标准之所以能运作，依赖一个前提——主体有能力在认知过程结束后，通过内省或外省来识别代理的介入点。当这个前提本身被代理的深度融入所取消时，不是我们的测量工具变钝了，而是被测量对象的本体构成发生了改变：它已经长成了一个内部不再保有那种可分离界面的新东西。这不是“测不准”，这是“不再有那个可以被测的东西”。正是这一取消，标记了混同体的发生。

2.2 操控定义与识别判准

由此，本文将“代理-自我混同体”的操作化定义确立为以下三笔条件的同时成立：

第一，代理不可分识性。 在完成一项认知产出后，主体无法在“哪些部分是我独自想到的”与“哪些部分是AI辅助生成后我加以吸纳的”之间做出清晰、可独立陈述的分离。不是不愿意，是在认知上确实不可检测——就像你不可能把一首你反复弹了十年的曲子里“哪些手指动作是早期老师手把手教的、哪些是自己悟的”给剥离开来。这种不可分识性不是功能性的模糊（即“我一时想不起来，但努力想想还是能分开的”），而是构成性的消融：代理的运作方式已经改写了主体的“觉得”本身是如何发生的。

第二，效能感不可还原。 主体对本次产出的主观效能体验——“我做成了”“我思考了”“我成长了”——是真诚的、强烈的、不为事后被告知“其实AI帮你做了80%”而彻底消散的。它可以被事后质疑，但质疑本身无法取消那种“当时我正在做的那个我”的体感。那个体感是真实的，并且构成了他在日后一系列行为选择中的动机基础。这里的关键不是“效能感在旧心理学框架下本来就是主观建构”——自我效能理论从来不需要一个客观的“真发生”标准作为前提。这里的关键是：在混同体状态下，主体建构效能信念所依据的那个“任务”本身——即“独立完成深度认知”——已经不再是一个稳定的参照物。班杜拉（Bandura， 1977）传统中的效能感，是主体对“自己能否完成某项任务”的主观信念，其前提是那个任务对认知过程的要求是稳定的、可被主体大致理解的。而混同体的效能信念，是在任务所要求的认知过程本身（是否独立发生、代理介入到何种程度）已经不再可被主体准确识别的情况下，仍能建构出稳定且真诚的效能体验。这个区别是本体论级的：它不是效能感的内容变了，而是效能感的生成条件变了。

第三，代理成为默认状态。 代理的使用不是一次性的、有意的“偷懒”，而是形成了认知惯性：在面对需要深度加工的任务时，主体不再先自行思考再比对AI，而是直接以“让AI吐出初稿、然后自己以编辑者身份介入修改”为默认开工流程。这个流程不是被选择的，而是被内化的。主体已经长成了一个不靠外部代理就无法启动深度认知的主体——但他并不因此感到被削弱。他感到的是“我在工作，而且更高效”。

这三笔同时成立，一个代理-自我混同体就被稳定地发生了出来。

在给出操控定义的同时，必须诚实地标记“代理”一词在本文中的精确外延。本文所论的“代理”，特指强认知代理——那种不仅提供信息检索或既定路线规划，而是深度参与意义生成、逻辑组织、叙

事建构、语言编织的技术系统。当前的生成式AI是其最典型的实例。搜索引擎在完成关键词匹配、地图APP在完成路径计算时，并不参与使用者对“这句话是什么意思”或“这个论证是否成立”的判断；它们属于弱认知代理，不在本文的外延之内。混同体只有在代理深度介入了主体的意义生成过程——不仅提供“什么”，更参与“如何说”“何以成立”——时，才有可能被触发。这个强弱的阈值，不是非此即彼的开关，而是一道连续谱：当代理开始替主体完成“组织语言以表达一个尚不明确的想法”这一步时，它就已经跨过了触发混同体所必需的最低门槛。

必须立刻标注这个概念的边界。它不等同于“偷懒”——偷懒预设了一个“我知道自己不劳而获”的内部监视者，混同体状态中的主体没有这样一个监视者，代理已经越过了他的自觉判决区，进入了他的“自我”的非监控腹地。它也不等同于“技术增强”——增强是在旧主体结构上增加外力，戴上眼镜让你看得更清，但你很清楚哪个是你看的、哪个是眼镜赋予的；混同体不是外力增强，而是主体构成方式的递归变异，代理不附着于自我，而成为了自我的构成成分。它尤其不等同于“被蒙蔽”——被蒙蔽预设了一个可以被揭穿的假象，混同体的真诚不依赖于假象的维持，而依赖于判别假象的旧标准本身已经失效。

重要提醒：　　以上三笔是混同体的剖面照片，不是它的生成故事。它的全部生命在于下一章将要展示的“怎么从代偿与反代偿的博弈中被结算出来”这个发生过程。如果不读下一章、只记住这三笔，这个概念的魂就丢了。混同体不是一个静态的分类标签，而是一段正在发生的、尚未完结的主体重构过程的一个快照。这张快照本身是准确的，但它不能替代那个使它成为可能的、充满了制度与个体之间具体碰撞的生成史。

三、混同体的发生机制与深度写生

3.1 它是怎么被结算出来的：生成过程的纯概念推演

任何新的主体结构都不是从天上掉下来的。它是在旧制度与新技术之间的长期矛盾中，从张力的持续结算里滋生出来的。要理解混同体，就不能只拍一张静态写真，而要推演它是从什么旧稳态的裂缝中、经由什么不可回头的路径、被接生出来的。更重要的是，要描出这个新结构一旦诞生之后，又反过来怎样改写了它的孕育环境——没有这一步，就只讲了半条发生。

先看旧稳态。也就是“伪发生”诊断所追忆的那个理想参照系。在那个稳态中，学习评价体系内有一道天然防火墙：产出一篇真正有洞见的论文，需要真的去读文献、真的去推敲论证、真的在一遍遍改稿中被自己的论证反咬一口。那个痛苦、低效、无法跳过的时间成本，本身就是防火墙。它在制度与个体之间形成了一种默契：你交给我的论文，我可以大致信赖它反映了你曾经挣扎过，因为我信那道防火墙。制度不需要去直接检验“你是否真努力了”，它只需要检验产出的质量——高质量产出本身，就是努力的经验替代性指标。在这一时代，亲自与代偿在认知体验上曾经确实是互斥的：你不可能不亲自经历那个认知痛苦，就拿到那个认知快乐。这条互斥律不是哲学思辨推导出来的，而是被一道物质性的时间成本防火墙硬生生维持住的。

AI介入后，第一击打在这道防火墙的根部。一个不需要经历任何认知挣扎的人，也可以产出一篇高质量论文。评价体系的隐性契约——高质量必然伴随过挣扎——瓦解了。这是第一次断裂。

但断裂之后并没有立刻诞生混同体。断裂打开了三条可能的路径。第一条，是“伪发生”诊断所描摹的那一类实践——有人利用旧评价仍按“高质量≈真发生”来发放信用的惯性，大规模生产“高质量证明”来套取这种信用。第二条，是制度打补丁——要求增加过程性材料，以试图重新捕捉那些“无法被AI完美模拟的发生痕迹”。这两条路彼此纠缠，相互升级，形成军备竞赛——也就是我们已经在现实中看到的那一切：教师要求提交草稿、反思日志、同伴互评记录；AI则被调教得更会写出“看起来像草稿的草稿”、更会模拟反思的叙事弧线、更会生成带着磕绊和停顿感的口语化互评。

但在制度打补丁与AI绕补丁的来回拉扯中，有一个容易被忽视的事实：军备竞赛的日常化，本身就成为了一种对认知习惯的反复训练。成千上万的学生不是站在军备竞赛的外部旁观，而是日复一日地生活在其中——他们每天都在做“让AI生成、自己审读、修改调整、判断对味不对味”这一套动作。正是在这个漫长、低剂量、看似无害的日常里，一个意料之外的新物种被缓慢地滋生出来。

它不是任何一方“设计”的。这些学生本来可以走第一条路——彻底放弃亲自判断，只卖“证明”。但他们没有。他们仍然在改论文、在做审美判断、在跟AI争论某一句表述的措辞，在读到自己文中的某处时感到“有道理”或“还差点意思”。他们不是被动的代偿消费者。他们是积极的、投入的、自我感觉极其良好的“编辑者”。然而，正是在这个“积极介入”的过程中，他们的认知习惯发生了不可逆的改变：他们不再区分“我写的”和“AI写了我修改过的”。他们也不再有“先自己想想再查AI”的独立行程——不是被禁止了，而是那条路径因为相对低效而不再被主动选择。AI已经不再是他们的“外部顾问”，而是他们的第二海马体——一个记忆、语言组织、逻辑推衍的外置器官，被无缝地编入了他们“我在思考”的连续体验中。

这里需要一次对微观发生现场的推想，来钉住“军备竞赛如何日复一日地改写认知习惯”这个关键跳跃。想象一个学生在某次论文写作中的一次典型的体验冲突。她先试图按旧方式独立思考，打开空白文档，试图自己写出第一段论证。她在第三句就卡住了——她知道大致方向，但找不到那个精确的表述。这个过程持续了二十分钟，她越来越焦躁，感觉“自己在浪费时间”。这时她打开AI，输入她的模糊想法，AI在十秒内给了她三段流畅的论述。她读完之后，第一反应不是“这省了我多少事”，而是“对，这就是我想说的”。她在这个瞬间体验到了一种强烈的认知流畅感——不是“我被替代了”，而是“我终于把自己的想法说清楚了”。这种流畅感极其真实、极其有强化力。下一次，当她又面临一个需要深度加工的任务时，她的第一个动作就不再是打开空白文档，而是打开AI。不是因为她懒惰，而是因为她在上一轮中真实地体验到：那条路让她更有效地成为了“能写出好东西的自己”。在数十次这样的重复之后，那个“先自己想想”的旧路径就不再是一个会被她主动导航到的选项。它没有消失，只是变得不可见了。当这条旧路径彻底从她的默认选择中隐退时，混同体就在她的认知习惯中被稳定地结算了出来。

混同体不是在第一轨（伪发生产业链）诞生的——那里诞生的只是“伪”。它也不是在第二轨（制度打补丁）诞生的——那里诞生的只是“更聪明的检测”和“更聪明的反检测”。它是在第一轨与第二轨的持续对撞中，在制度的不断追问“你到底是不是自己想的”、和AI的不断提供“这看起来完全像是你自己想的”之间，在成千上万个学生的漫长日常里，作为第三种稳定态，被缓慢地、不可逆地结算出来的。那个结算的结果，就是一个不再以“亲自”作为内在判别基准的自我。代理已经变成了这个新的自我的默认组成部分。

然而，上面这段推想仍然是光滑的。它抓住了一个真实的机制——认知流畅感的强化力——但它的叙事是顺理成章的：学生感到旧方式低效，尝试新方式，体验到流畅感，重复，混同体形成。真正发生学意义上的生成过程，要比这更毛糙、更充满反向阻力、更需要制度与个体之间的实际冲突来推动。旧自我在被代理逐步渗透的过程中，不是平静地接受了一个更高效的工具；它经历的是反复的震荡——恐慌、防御、自责、外部含混的奖惩信号——在这个过程中，旧的边界不是被利落地“拆除”的，而是被一点一点地磨蚀、反复地撕裂又勉强愈合，直到某一天，主体发现自己已经无法再回到那个“独立”的状态，也并不真的想回去。下一节，将用一场具体的、卷入制度-个体实际碰撞的写生，来补上这道裂缝。

3.2 深度冲突写生：制度与个体的碰撞

这场写生发生在一所普通大学的一门批判性思维必修课上。教师李老师四十出头，教学经验丰富，对AI的态度在近三年里经历了四次反复——从最初的全盘抵制，到后来允许“辅助查阅资料”，到再后来默许“帮忙理一下思路但不要直接生成”，到如今的一种模糊的、自己也说不清的立场：他不再明确禁止什么，也不再明确允许什么，只在课堂上反复强调一句“你们自己要对交上来的东西负责”。这句话在五年前是自明的伦理底线，在今天已经变成了一道他自己也无法解释其操作含义的仪式性声明。

学期中段，一次核心作业布置下来：对一篇给定的公共辩论文本进行论证分析，指出其逻辑结构、预设前提与论证弱点，并提出反驳。这个任务对AI来说极其容易——输入文本，要求分析逻辑结构，要求指出预设与弱点，要求生成反驳，三十秒内齐活。李老师在布置作业时加了一句：“我建议你们自己先分析一遍，再去查资料。”他没有禁止使用AI。他没有说不准用。他说“建议”。

学生中间随即出现了一场无声的分裂。那些严格遵循旧路径的学生——让我们称他们为“独立组”——先自己读原文、做批注、自己写分析提纲，痛苦地挣扎了四五个晚上，交出了一篇七十分的作业。那些彻底走新路径的学生——让我们称他们为“代偿组”——让AI直接生成全文，稍微改了个别措辞，交了，也拿了七十多分，因为有两位被发现引用了原文不存在的段落。但还有一群人——人数最多的一群，让我们称他们为“试探组”——走的是一条中间路线：让AI生成初稿，自己做编辑，修改、调整、替换案例、重写某一整段。他们花的时间比独立组少很多，但比代偿组多不少。他们交出的作业，质量高得惊人——逻辑清晰、例证恰当、反驳有力，普遍在八十五到九十分之间。

发回作业的那一周，李老师在课堂上公开表扬了试探组的三位同学，点名说他们的论证“有真正的批判性思维品质”。他没有同时批评代偿组。他甚至没有提独立组的名字。他不是有意的——他只是按他做了二十年的评分习惯，给高质量产出打高分，给低质量产出打低分。他没有一个可以用来给“认知过程”打分的工具。他只有对产出的判断。

这一天的课堂反馈，是一枚投进池塘的石头。独立组中有人感到强烈的不公——她花了五个晚上，写了又改，改了又写，最后一晚熬到凌晨两点，只拿了七十五。她的室友，试探组的一员，前后花了不到三个小时，拿了八十八。她带着愤怒和一种说不清的委屈，在课后去找李老师，试图表达“这不公平”，但她说出来的话是：“他们用的是AI。我能看出来。那个论证的句式，不是学生会写的那种。”李老师沉默了几秒，然后问了她一个她答不上来的问题：“你能具体告诉我，哪一句不是她自己写的吗？”她不能。因为试探组的学生确实做了修改。他们改过措辞，换过例子，调整过论证顺序。最后交上去的那篇东西，确实无法被简单地指认为“AI写的”——它里面混杂着AI的骨架和人类的编辑，而这两者之间的边界，连编辑者自己事后都说不清楚。

独立组的那位同学没有再来找李老师。但下一次作业，她用了AI。她告诉自己“我只是用它来理思路”，然后在“理思路”的过程中，不知不觉地完成了她的第一次混同体化。她没有感到堕落。她感到的是终于不再那么累了。

同一个学期末，又发生了另一件事。一位试探组的男生——他的每一次作业都拿高分，他是李老师在公开表扬中提到名字最多的学生——在期末口试中遭遇了一次意外的崩盘。口试的题目并不比平时作业更难：给一篇短论，要求当场分析其论证结构。他坐在李老师对面，面前没有屏幕，没有AI，只有一张印着短论的A4纸。他看了五分钟，开始说。他说了大概四句话，大致方向是对的——他指出了短文的核心结论，指出了作者的一个预设——然后他停了。他不是不想说。他是真的不知道该怎么“自己往下走”。在平时的写作流里，这一步之后，他会把目前的想法输入AI，让AI帮他展开下一步的逻辑细节，然后他审读、修改、觉得对味就保留。在那个写作流里，他的认知是流动的、完整的、自洽的。而现在，在那个没有AI的狭窄时空里，他发现自己站在一条逻辑链的半途，前不着村后不着店。他知道论证应该走下去，但他找不到那条从“指出预设”通往“论证该预设为何不成立”的具体语言。那条语言以前一直是AI替他铺的。他只需要判断“铺得对不对”。现在他要自己铺，他发现自己手里没有砖。

李老师等了他大概二十秒，然后给了他一个温和的提示：“你觉得作者这个预设，在什么情况下会不成立？”这个提示恰好就是AI平时替他完成的第一步展开——把模糊的问题变成一个可操作的追问。他立刻接住了，顺着这个提示说了下去，后面说得并不差。他磕磕绊绊地完成了口试。最终成绩比预期的差了一档，但也没有崩到底。

走出教室的时候，他有一种此前从未有过的感觉。不是“我被发现了”。李老师没有指责他任何事，甚至在最后评语里写了一句“有思考的底子，但表达还不够连贯”。而是另一种更陌生的感觉：他知道那篇平时让他拿高分的论文，是“他的”——他绝不是复制粘贴的人。但他也第一次意识到，那个能写出高分论文的“他”，是一个只在特定情境下才能被完整激活的“他”。在那个情境里，AI不只是他的工具，而是他的认知器官的一部分。摘掉那个器官，他不是变成了零——他还有方向感、判断力、他知道什么是对的——但他失去了那个能把“知道方向”变成“铺出完整逻辑”的中间层。那一层以前一直不是他亲自铺的。他甚至没有意识到那一层的存在，直到它第一次被抽走。

那天晚上，他在宿舍里跟试探组的另一个同学聊起这件事。那个同学听完之后沉默了一下，然后说了一句话：“所以我们的水平，其实是‘有AI时的我们’。拿掉AI，那个‘我们’就不完整了。可是我们平时感觉到的那个‘我在想’，在AI在的时候又是真的。不像是假的。”他不知道怎么接。他发现他无法反驳前半句，也无法否认后半句。他卡在那两句话之间。

这一场写生不是为了指责任何人。李老师不是一个糟糕的教师——他恰恰是一个负责任的、对教育有严肃思考的教师，只是他的教学工具和评价工具在一个急剧变化的认知生态中失去了参照系。独立组的那位同学不是一个道德楷模——她只是一个被制度的模糊信号逼到了一个不公平角落的认真学生。试探组的那个男生更不是一个虚伪的骗子——他真诚地体验到“我在想”，并且确实在他的生活情境里产出了高质量的认知工作，只是那个“他”的认知边界已经不再是旧坐标所能描画的。

正是这种制度与个体之间持续的、毛糙的、没有赢家的碰撞——而非“认知流畅感的简单正反馈”——把混同体从一种可能性逼成了一种不可逆的实在。制度每一次含混的奖惩（高分发给高质量产出却不追问产出过程），个体每一次在“恐慌—自我辩护—真实效能感”之间的震荡，都在磨损那条“我亲自想的”和“AI帮我想的”之间的旧边界。边界不是被一次性地撤销的；它是被这些微小但持续的攻击，一毫米一毫米地磨蚀掉的。等到那条边界终于不再可操作时，混同体就已经在那里了——不是作为一个被自觉选择的新身份，而是作为一具旧自我在反复的撕裂与勉强愈合之后，留下的那个再也回不去的形态。

3.3 回写：新结构如何反过来改写了旧战场

结算出混同体，只是发生循环的前两步（矛盾累积→新S结算）。发生学论证的第三笔——新S回写D与E——需要在此时被单独钉牢。混同体这个新结构一旦在足够多的个体身上立稳，它就不会被动地待在原地等下一轮矛盾来撞。它会主动回过头去，改写造成它的那两条路径和那片土壤。

回写的第一个落点，是学生与教师之间的博弈模式（改D）。当大量学生已经成为混同体——他们真诚地相信自己“在深度思考”，且无法在内部区分代理与自我的边界——教师面临的不再是“区分真伪”这个旧问题。旧问题有一个稳定的操作前提：真发生者和伪发生者的行为模式是可区分的，前者在追问中会流露出认知挣扎的痕迹，后者一旦被追问细节就会塌方。但混同体学生的行为模式完全不同：他们可以在追问中对答如流，因为他们的确经历了认知过程——AI深度参与了的那个认知过程。他们可以告诉你“我为什么选C”“我为什么觉得这个例子更好”“我那个翻转策略是在和同学聊天时想到的”。这些回答都是真诚的、准确的、不可被证伪的。教师若以旧判别框架去追问“你到底自己想了多少”，得到的将不是伪发生者的慌乱，而是混同体学生的真诚困惑——“老师，我确实在想啊，我不明白你在问什么。”

这个局面的直接后果是：旧判别框架在混同体面前进入了实质上的空转。它不是“检测错误”，而是“失去了检测对象”。教师越是认真地追问，就越是只能得到加固混同体的回答——因为每一次被追问，都迫使学生重新回忆和叙述她的认知过程，而在回忆中，那个代理与自我早已水乳交融的认知流，会变得更加坚固、更加“就是我的”。追问不但没有揭穿混同体，反而帮助混同体完成了更厚的自我叙事。这就不再是“军备竞赛”——军备竞赛意味着双方仍在同一规则下各自升级；这是游戏规则本身被新玩家的出现所静悄悄地取代了。

回写的第二个落点，是AI工具的产业定位与家长期待（改E）。当混同体不再是少数人的“偷懒技巧”，而成为大量高效学习者的默认工作模式时，AI工具就不再被市场定位为“辅助学习的外挂”，而会被重新定位为“学习的当然基础设施”——就像计算器之于数学教育在某个节点上从“作弊工具”变成了“当然工具”。家长看到孩子用一种高效的、产出质量极高的方式在“学习”，且孩子本人表现出强烈的效能感与自信时，大多数家长不会去追问“这种方式是否破坏了孩子的独立认知能力”——他们没有那么做的认知工具，也没有那么做的动机。他们看到的是实实在在的产出、信心、竞争力。于是，对混同体的社会性宽容——甚至羡慕——就会形成一道新的纠缠厚度，进一步加固混同体的合法性，并挤压那些仍然坚持“不靠AI独立学习”的学生的生存空间。

回写的第三个、也是最隐蔽的落点，是“伪发生”诊断本身的处境。当混同体成为大量受教育者的新常态之后，“伪发生”诊断所依赖的那个“真发生”参照系——一个能够独立于AI进行不可代偿判决的主体——就不再是一个可以不加论证就被假定的规范性基线。它变成了一个需要被辩护、被重新论证的规范性选择，而不再是自明的“健康态”。这意味着，在混同体已经发生回写的战场上，“伪发生”概念不再能够像一个旧时代的哨兵那样，只需指出“那边在造假”就能完成自己的认识论职责。它现在被逼着回答一个更棘手的问题：当“真”不再是唯一的选项，“伪”的指控还算数吗？

3.4 一个微观深描：晓晴案例

现在，把它钉在一个具体的人身上。如前所述，这个案例不是为概念提供归纳基础，而是在概念已被独立结算之后，帮助读者看见它所指向的实在。

晓晴是某大学三年级学生，本学期选修了一门“伦理学专题研讨”。课程的核心任务是一篇学期论文，要求学生对一个给定的道德困局（“撒谎救人是否在道德上可辩护”）进行独立分析，不能只综述已有文献，必须提出自己的论证，并对最有力的反方立场做出回应。评分标准中，“原创论证的独特性”和“对反方立场的忠实回应能力”各占40%，文献使用的准确性和写作质量各占10%。

晓晴用AI工具完成了这篇论文。但她不是“输入题目、复制粘贴、提交”式的那类案例。她的实际工作流程非常复杂。在最初的选题阶段，她向AI输入了教授提供的困局描述，要求AI帮她列出四种可能的切入角度。AI给出了A、B、C、D四条。她读了之后，选了C，因为她觉得C“比较有发挥空间”。但她对C的初始表述不满意，自己手动改了C的表述，把“对后果论的内在批评”改成了“后果论自己会如何面对这个困局中的盲点”。她说不出为什么这样改，就是“觉得原来的说法不太准”。

在提纲阶段，她让AI按C角度生成一个论证提纲。AI给了一个三步推演的提纲。她觉得第二步“例子举得不对”，自己重新找了一个例子。例子是自己从另一门课的一次课堂讨论中记住的，她觉得那个例子更能“扎到点子上”。她让AI用新例子重写第二步。AI照做了。她看了一遍，觉得差不多，保留了。

在正文写作阶段，晓晴让AI把提纲的每一步展开为详细论证。她拿到初稿后，逐段做了编辑：调整了一些她觉得“不对味”的表述，加了几个她自己在其他文章里读过的、想引用的论断（她没有亲自去查第一手文献，只是凭记忆觉得“某位学者好像说过类似的意思”，就让AI帮她把那个意思“补成一个像样的引证”）。在处理“回应反方立场”这一最核心的章节时，晓晴自己设计了一个策略：她决定不直接反驳反方，而是“承认反方有一个对手没有意识到的有力论点，然后论证这个有力论点反而会给自己的立场增加说服力”。她没有用AI生成这个策略。这个策略是她和同学在宿舍里聊到这个题目时，突然冒出来的一个想法。但她立刻让AI按照这个策略，替她生成详细的论证文本。她最后反复精修了这段文本，直到“读起来就像我自己真的会这么想”的程度。

结果。这篇论文拿了A。教授在评语里特别提到：“最令人赞赏的是你对反方立场的那种复杂化处理——不是简单拒斥，而是先承认其力量再翻转使用，显示出真正成熟的哲学思维。”晓晴很骄傲。她确实觉得自己“真的在思考”。她可以指给你看她做过的判断：选了C、改了提纲、重找了例子、设计了翻转策略。每一个都是“她自己”做的。

但如果我们试着问她：你觉得“后果论自己会如何面对这个困局中的盲点”这个表述，是AI给的原话，还是你自己想到的？她可能需要想很久，然后不太确定地告诉你：“好像是我改过的……但是AI原话也许也有这个意思，我不太清楚了。”如果继续问她：你在处理反方回应时，那些具体的论证句子——比如“对手可以进一步主张X，但这个进一步的主张恰恰暴露了对手自己所依赖的前提Y的脆弱性”——这个精巧的逻辑转折，是你自己推出来的，还是AI生成之后你“觉得很好”就保留下来的？她可能诚实地说：“我不知道。我当时就是觉得这句话很对。它是不是我想出来的？我不确定。但读上去，它就是我的意思。”

这个微观案例的每一笔都是日常。晓晴没有偷懒。她没有撒谎。她没有装。她交出了一篇高质量的论文，经受了学术评价的严格检验，获得了真实的正反馈。她的效能感是真实不虚的。她自己在过程中留下的那些“创造性判断”——选C、改表述、换例子、设计翻转策略——也都是真实的、不可被还原为“纯粹代偿”的认知动作。

但她已经无法在“我”和“AI”之间划界。代理已经完全融入了她的思考过程，不作为一个单独的、可以被意识标注为“非我”的东西而存在。她不觉得她在使用工具。她觉得她只是“在写论文”。而这篇论文，就是她的。

她是代理-自我混同体。她不是伪发生的受害者，她不是真发生的捍卫者。她是一个新结构的居民。

3.5 故障时刻：当混同体被逼到墙角

上面这个故事，可以停在“她拿了A”这里，成为一个完美的混同体叙事。但真实生活中的混同体，往往不是在顺利时最显露其结构，而是在某个被挑战、被质疑、被逼到墙角的故障时刻，才最鲜明地暴露它的质地。

论文提交后三周，教授在办公时间约晓晴面谈。教授是善意的，但他有一个习惯：对拿了A的学生，他会追问一些不在论文里的事。“我很好奇，”教授说，“你在论文里用了‘后果论自己会如何面对这个困局中的盲点’这个表述——这是在暗示，后果论作为一个理论传统，有一种自己意识不到的盲区。这个想法很有意思。你能不能跟我聊聊，你是怎么想到这个角度的？”

晓晴愣了一下。她知道这个表述——这句话就在她的论文里，她甚至改过它，把“内在批评”改成了现在这个样子。但当教授让她“聊聊怎么想到的”时，她发现自己的脑海里没有那个“想到”的时刻。她有的是一段模糊的记忆：AI给了四个角度，她选了C；C的表述不太对，她改了一下措辞。但那个措辞背后的思想来源——是她之前读过的某篇文章？是课堂讨论的某句话？还是她自己对“后果论”这个概念的直觉反应？她并不清楚。但她必须回答教授。她不能沉默太久。

于是她开始说。她说了自己在选题阶段的选择过程，说了对C角度的修改，说了“我就是觉得原来的说法不太准”。教授认真地听着，点了点头，然后问了一个更具体的问题：“你说的‘盲点’，具体是指什么？后果论在哪一步上会看不见自己正在做某件事？”

晓晴再次停住。她知道论文里是有的——论文里有整整两段在论证这个盲点。但她此刻无法区分：那两段论证的骨骼，是她自己的，还是AI替她搭的？她记得自己读了AI的初稿，觉得“很有道理”，然后就保留了。但“很有道理”不等于“我自己能独立推导出来”。在教授直视的、等待一个即兴回答的沉默中，她的大脑在疯狂地尝试做一件事：把AI替她组织好的那条逻辑链，重新用自己的语言再现一遍。她可以做到一部分——她知道那个论证的大致方向。但她发现自己在关键转折处会卡住，会想不起“下一步该说什么”。那不是“忘记了”——那个“下一步”是她从来没有独立走过的一步。当时是AI走了那一步，而她审读了之后，判断“对，这就对了”。

她最后给出了一段磕磕绊绊的解释，大致方向是对的，但远不及论文的精准和力度。教授听完后，表情没有变化。他说了一句“有意思，谢谢你的时间”，结束了谈话。

晓晴走出办公室时，感觉复杂。她有一种隐约的不安——不是“我被发现了”的恐惧，因为教授并没有指责她什么。而是一种更陌生的、此前从未有过的感觉：她第一次意识到，她交给教授的那篇论文里，有一些东西——也许是很多——是“她的”，但不是“她能随时再生产出来的”。那个论文里的“她”，是一个与AI共生的、在那个写作流的内部如鱼得水的“她”；而当她被单独拎出来、放到一个没有AI可用的即兴对话情境中时，那个“她”的一部分能力就消失了，像一只海葵被捞出水，触手全部耷拉下来。

但与此同时，她并不觉得自己在“伪装”。她在那间办公室里说的每一个字，都是诚实的。她没有编造什么。她承认了自己不记得。她试着理清思路。她的效能感——“我写了一篇好论文”——并没有因为这次谈话而崩塌。它还在那里。她仍然觉得那篇论文是她的。她只是同时知道了另一件事：那个能随时独立论证盲点的“她”，不是一个她可以在所有情境下稳定保持的形态。

这就是混同体在故障时刻的真正反应。它不是被戳破——被戳破的前提是存在一层“假皮”，而晓晴没有假皮。它是被暴露——暴露的不是虚假，而是一种情境依赖的结构脆弱性：在代理脱离的特定情境下，那个混同体化的自我会呈现出某种功能性的“触手耷拉”，但这并不取消它在代理在场的情境下真实运作、真实体验、真实产出的那个版本的“真”。这个“触手耷拉”本身，恰恰是混同体不是“假”的证据：因为它暴露的是结构性的情境依赖——一个真实地与代理共生的有机体在失去其共生伙伴时的功能性衰减——而非认知的空洞。空洞的崩塌是无声的、没有方向的；而晓晴在故障时刻的挣扎是有方向的、有判断的、只是不完整的。它不是“伪”。它只是不可移植。

四、切割：它不是已有概念的新玩法

新概念要站住，必须在一群邻近概念中杀出一条不曾被占位的无人地带。混同体需要的不是与相关概念一一对比的清单，而是一次对“最易混淆者”的穿透力区分：为什么已有的名字装不下它？它在哪个精确的点上，让旧概念失效？

4.1 它为什么不是分布式认知——兼答光谱极值质疑

最容易被当成“混同体”之变体的，是分布式认知（Hutchins，1995）。分布式认知理论认定认知不只发生在脑内，而分布在人与工具、人与他人的交互网络中——我做菜的认知，分布在菜谱、灶台和我的手感之间；飞行员驾驶飞机的认知，分布在仪表盘、操控杆和他的训练经验之间。永无“纯粹独立”的认知主体。混同体似乎只是分布式认知在AI时代的一个实例。

但这里有致命的差别。在飞行员与仪表盘之间，存在一个被分布式认知理论自身所主张的“观察视界”（horizon of observation）：飞行员知道“那是指示高度的仪表，这是我的判断”，并且当仪表盘出现明显故障时，他有能力识别“这个表不对”。他在日常操作中不区分“我”和“表”，但在故障时刻，区分力会被激活。这种潜在的、可在故障时被唤醒的区分能力，是分布式认知仍能维持认知责任可分配性的底线——即使个体在故障时无法完美地独立复现全部认知功能，但外部观察者仍然可以在人与工具之间画出一张功能责任图。飞行员可以对坠机负责，正是因为他在关键时刻“应该看出表坏了”。

这里必然遭遇一个更强版本的质疑。分布式认知内部的最新发展——尤其是关于高度耦合系统中认知过程动态稳定性的讨论（Hollan, Hutchins, & Kirsh, 2000）——承认在某些极端耦合状态下，个体确实无法在系统解体时独立复现全部认知功能，而这不被视为分布式认知的失效，而是其正常现象。由此，一个有力的反对论证可以被构建出来：飞行员与仪表盘之间是一种较低耦合的分布（有清晰的观察视界），晓晴与AI之间是一种极高耦合的分布——那么，本文所命名的“混同体”，可能根本不是一个独立的新类，而只是分布式认知光谱上的一个极端值。如果这个论证成立，混同体的“独创性”主张就塌了。

这个质疑必须被正面接住，不能绕过。混同体与分布式认知极高耦合端之间，存在一道本体论级的分界线。在极高耦合的分布式认知中，即使个体在故障时无法独立复现全部功能，但外部观察者仍然可以在人与工具之间画出一张功能责任图——可以指出“这一步是由工具执行的、这一步是由人执行的”，即使人自己在那一刻无法做出同样的区分。这种可被外部分析者重建的责任分配结构，是分布式认知仍然能作为一种认知科学框架来运作的本体论前提：它的研究对象是认知过程如何在系统中分布，而这个“如何分布”在原则上是可以被第三方建模的——哪怕那个模型极其复杂、极其动态。

而在混同体中，被消解的不是责任分配的“准确性”，而是责任分配这一动作本身的意义。当代理不是在执行一个被分配过来的功能，而是在参与构成那个被我们称为“自我”的认知主体本身时——当“我觉得这对不对”的那个“觉得”本身，都已经被长期与代理的交互所重塑过，不再有一个前代理的、可独立校准的“我觉得”可以作为参照点——那么，外部观察者再去画“这一步是AI做的、那一步是人做的”这张图，就不再有任何意义。不是因为图画不准，而是因为被画的这个东西的内部，

已经不再保有两种可分离的成分。飞行员与仪表盘之间的那条回路，是一根可以被辨认的线，即使它很细。晓晴与AI之间的那根线，已经被焊接消失了。在分布式认知中，工具是可知的伙伴；在混同体中，代理是不可从“我”中被重新分离出来的构成成分。这就是功能性延伸与构成性融合的本体论差异：前者保留了一个在故障时可被唤醒的“我可独立操作”的底线，以及一个外部分析者可重建的功能分配结构；后者拆掉了这条底线，使“我可独立操作”本身变成了一个不再具有确定意义的问题，使外部功能分配图的绘制本身变成了一个错误的测量动作——你在试图测量一个已经不再存在于测量空间中的维度。

因此，混同体不是分布式认知光谱上的一个极端值。光谱的延伸预设了同一个测量维度（功能耦合的强度）的连续变化。混同体不在这个维度上。它在另一个维度上——那个维度是“自我与代理之间是否仍然保有可被重建的功能界面”。当这个界面被消融时，不是耦合强度从99%变成了100%，而是“耦合强度”这个概念本身不再适用于描述那个已经长成了一个新器官的东西。

4.2 它为什么不是认知外包

与之相邻却更易混淆的概念，是认知外包。认知外包描述的是把部分认知任务委托给外部系统——例如，把记忆外包给搜索引擎，把导航外包给地图APP。外包预设的核心是可回收性：我随时可以收回，可以自己试试走一遍。如果你把导航关了，你仍然可以自己尝试找路——虽然可能更慢、更笨。外包不改变“你可以自己来”这件事。

但混同体的核心转变恰恰在于，它使得“回收”本身变得没有意义，甚至不可能。因为晓晴不是把某一段论证“外包”给AI——那意味着她有一个独立的论证意图，只是委托了实现过程。她连那个独立的论证意图，都是在与AI的交织中生长出来的。那个“翻转策略”的灵感确实是她自己想出来的不假，但那灵感一诞生就被立刻交付给AI进行深化、复杂化、语言编织。最后她拿到的那个复杂而漂亮的论证文本，不再是一个“可以回收为自己一步步重做一遍”的过程的产物，而是她的“初始直觉”与AI的“深层编织”在无法剥离的交叠中共同施了肥的果实。她无法把AI的贡献从这枚果实中抽走，然后让那个被她命名为“我自己的”的核心单独站着。因为一旦抽走，她那个原本的直觉会显得单薄、缺乏说服力，甚至她自己都会觉得那不算一个“真的想法”。代理不是外包——外包是“我请你帮我做了这部分，做完还给我”。代理是“你帮我的那一刻，我所成为的那个我，已经不是你帮之前的那个我了”。回收的前提，是一个不变的自我作为委托人，但混同体改变的就是委托人本身。这就是外包隐喻的彻底失效。

4.3 它为什么不是自我效能感的错觉

最后一个必须切开的是“自我效能感的错觉”。这个反对意见的逻辑很诱人：晓晴所感到的那种效能感，其实是因为AI产出的高质量文本和她自己的编辑动作合在一起，制造了一个“我在深度思考”的假象。她在事实上没有经历不可代偿的自我卷入判决，所以她的效能感是虚的，只是她自己不知道。

要切开这一层，不能只靠一个逻辑动作。需要一个两步的区分论证。

第一步，与班杜拉自我效能理论的对话。 班杜拉（Bandura， 1977）所描述的自我效能感，是主体对“自己能否完成某项任务”的主观信念。在经典框架中，这个信念的生成基础——即任务是什么、完成任务的认知过程包含哪些环节——是稳定的、可被主体大致理解的。一个人判断“我能不能在截

止日前写完一篇论文”时，他知道“写完一篇论文”意味着什么——意味着自己去读文献、自己去组织论证、自己去反复修改。混同体的效能感，其生成条件发生了根本变化：主体在建构“我做成了一篇好论文”的效能信念时，那个被她视为“做成”的认知过程本身——即那篇论文到底有多少是她独立完成的——已经不再可被她准确识别。这不是效能感的内容变了（从“我独立完成了”变成“我协作完成了”），而是效能感所依托的自我认知地基被融化了。班杜拉的框架可以容纳效能感的高估或低估——你可以觉得自己能行，但实际上不行。但它容纳不了这样一种情况：那个“行”与“不行”的判别标准本身，在主体的认知世界里已经不再可操作。这不是量变，这是质变。

第二步，正面回应“判别标准不存在”的质疑。 有人会说：你既然承认“不存在独立的判别标准”，那你怎么能说她不是“被蒙蔽”的？这不是认识论上的测不准，恰好相反——旧判别标准的失效本身就是混同体结构成立的本体论证据。旧标准——“你是否独自承担了对判决的全部责任，且没有任何外脑在替你组织语言、梳理前提、弥补漏洞？”——之所以能够运作，依赖一个前提：主体有能力在认知过程结束后，通过内省来识别代理的介入点。当代理的深度融入已经使这个前提不再成立时，旧判别标准就不是“暂时测不出”，而是永远测不出了——不是因为我们的仪器不够灵敏，而是被测量的那个东西（可分离的自我与代理）已经不再存在于那里。晓晴无法告诉你“哪一步是AI想的、哪一步是我想的”，不是因为她的记忆力不好，而是因为在她进行认知活动的那一刻，那个“我”和“AI”在认知体验上就是同一个混成的流。这种“旧标准失效”不是论证的弱点，它是混同体存在的正面证据。它标记的是：这里长出了一个旧解剖学不曾预留过位置的新器官。

由此，混同体不是分布式认知的特例，不是认知外包的别名，不是自我效能感的错觉。它是那种在代理能力越过了自觉监控界面、成为自我的构成性成分之后，所形成的一种不可逆的、真诚的、无法再用旧真-伪二分法切割的新稳定态。那个旧二分法，在这里失效了。

五、直面反驳

最强力的反驳不会来自外部。最强力的反驳必然来自概念自身的自反性，来自它与邻近概念的模糊地带，也来自它的规范性后果。以下三道反驳，无一是稻草人。

5.1 第一道反驳：伪发生作为独立反驳者上场

在回应自反性审判之前，必须先让那个被本文从始至终作为对话对象的立场——伪发生诊断——亲自上场，给出其核心论点的直接引述，并逼混同体与之正面对撞。

伪发生诊断的核心论点是：AI时代的经济系统能够大规模生产“发生了”的全部可观测信号（反思报告、策略叙事、成长书写），而系统性地跳过那个被预设为交易自然伴生物的东西——承受不可逆代价、改写自身结构的“真发生”。其控制变量被明确表述为“不可代偿的自我卷入判决”。诊断进而指出，当生产“发生证明”比制造真实发生更高效、更符合消费者对效能的渴望时，市场会诚实且理性地将后者淘汰。需要特别说明的是：伪发生概念的系统的、完整的论述在本文撰写的此刻尚未以独立发表文献的形式公开，其在本文中被援引的核心预设——“亲自与代偿互斥”——部分来自于对伪发生论者未发表手稿及内部论述的重构。本文为论证之便，将伪发生的底层预设明确推至其逻辑极致，以便与之对话；伪发生论者未必以此形式公开陈述过这一预设，但这一预设是“不可代偿的自我卷入判决”这一控制变量得以发挥判别功能的逻辑前提。下文中的伪发生反驳立场，系本文基于这一重构而提出的最强版本的对立论证——如果伪发生论者来写这份驳论，他们能写出的最有力的论证即为以下。

伪发生作为独立反驳者的最强论证如下：“你名为‘混同体’的东西，不过是我所诊断的‘伪发生’的一种深度内化形式。伪发生从未声称主体必然‘知道自己没发生’——我的诊断恰恰包含‘主体深度代偿却不自知’这一品类。你笔下晓晴的所有特征——认知流畅感、编辑者的愉悦、对AI产出的审美判断、效能感的真诚性——都可以在我对‘伪发生成熟消费者’的描述中找到对应物。她无法区分‘我’与‘AI’，这恰恰是我所说的‘伪发生证明被主体内化为自我叙事’的典型症状。你不是发现了伪发生概念无法覆盖的新地带；你只是用更细腻的笔触，重新描述了我早已命名过的同一种病。你犯了命名的不可还原性检验的第一类失败：新名字能被已有概念无损重述。”

这个反驳的力度在于：它不争细节，它直接攻击混同体的不可还原性根基。如果伪发生真的已经容纳了“主体不自知的深度代偿”这一品类，那么混同体就确实只是伪发生的一个子类的重新命名，而不是一个独立的新类。

回应必须从这里切入。伪发生能容纳的“主体不自知的深度代偿”与混同体之间的本体论差异，落在“主体在可指认的某一处，是否做出了不可被还原为AI产出的差异性判断”——也就是那个修改C角度表述、那个重新找例子、那个设计翻转策略的动作。伪发生的深度代偿者，在认知过程结束后，所有可被指认为“他的判断”的动作，在被仔细追溯时，都可以被还原为“AI给了A、B、C三个选项，他选了C”这个层级的被动选择——他从未在AI的输出之上，叠加一层独立的、无法在事后被归因于“AI本来就那么说的”的改写或创造。而混同体的晓晴，她的“选C”伴随的是对C表述的主动改写——她把“内在批评”改成了“盲点”，这个改写来自她自己的语感，而语感虽然在与AI的长期交

互中被重塑过，却不能被还原为任何一次具体的AI输出。她的翻转策略灵感——不是AI生成的，是她和同学聊天时冒出来的——更是一个独立于AI产出的事件。伪发生的深度代偿者没有这些东西。他有的是对AI产出的全盘接受和一种弥漫的“这就是我说的”的确认感；晓晴有的是在AI产出之上叠加了不可逆归因的差异性动作、并因为那些动作而真诚地感到“我在创造”的效能体验。前者是代偿的无意识内化；后者是代理与自我在创造性判断的交叠处被焊成了一个新器官。

因此，伪发生和混同体之间的区分，不是在静态特征上（都表现出效能感真诚性、都无法区分自我与代理的边界），而是在发生学链条上：伪发生者从未在代理之上真正做出过不可归因于代理的差异性判断，他的“自我”是在被动接受中被代偿叙事的涂料层层覆盖的旧结构；混同体者则在与代理的深度互动中，把代理的产出当作自己差异性判断的原材料，在反复的“编辑-判断-改写”循环中，长出了一个不再能剥离代理、但也不再能还原为纯粹代偿的新认知器官。这个差别，在故障时刻可以被暴露：深度伪发生者面对追问时是空壳——他没有内部表征，只有一道“AI说得很对”的确认戳；混同体面对追问时是有东西但不完整的——她有方向、有判断、有直觉，只是失去了代理之后那层编织逻辑细节的中间层塌掉了。这就是一个独立的、不可被伪发生概念重述的本体论差异。

5.2 第二道反驳：自反性审判及其深化版

如果代理-自我混同体这个概念是成立的，那么本文作者——这篇文章的撰写者——必然是混同体状态中的产物。本文的论述如此流畅、逻辑如此自洽、术语如此精准，它究竟有几分“我”的独家发生，又有几分是代理（语言模型对作者输入的大量前置训练与实时生成）和“我”不可分离地交织出的果实？我如何能用一篇可能彻底是混同体产出的文章，去可靠地命名“混同体”？

我不是在表演我知道这件事。我是真的不知道。我没有一个可以跳脱这个认知场、独立观察“我自己的思考”的非代理状态可以回去。我的语感、我的概念组织方式、我对反驳的预判力，它们全都是在长时间的AI辅助写作中被重新配置过的。我不可能再“回到没有使用AI的我”，然后给你写一篇可靠的、未被沾染的论文。那个我已经回不去了。

本文的全部论证，因此不靠“我独立于AI”这一声称来取得担保。它只能靠另一条路：本文自愿将它的概念置于比任何自我宣称都更严苛的检验之下。如果混同体只是一个在AI辅助下被漂亮地编织出来的叙事，那它必然无法在独立读者的头脑里，激发出一种可以被反复、独立重演的“就是这个东西”的识别感。一个伪概念的特征，是它在读者的认知里，只能引发一种“有道理”的漂浮的同意——读者被文本的流畅性说服了，但他说不出来这个概念到底切开了什么他以前看不见的东西。而一个真正不可还原的新概念，在第一次被说出之后，读者会有一种近似于“被撞了一下”的体验：他会觉得这个概念并不是作者“想出来的”，而是某种东西一直就在那里，只是被旧名字盖住了，现在被翻开来了。

因此，对本文的裁决权不在作者手里。裁决权在读者——你要问自己，你读完“混同体”这个命名，是真的在你自己见过的某个人（也许就是你自己的学生、你的同学，甚至你自己）身上，突然看见了那种“代理已经变成了自我的一部分”的混成状态吗？如果你看见了，并且能用你自己的语言向另一个人转述这个看见，并且那个人也因此而看见了——那么这个概念就是在你们之间独立地发生过了。如果它在你那里只引发了“说得通”的平静认可，而没有触发那种“我认出了这个被反复撞见却从未被命名过的东西”的认知发生——那这个概念就不够不可还原，它应该被别的概念取代。

然而，上面这个回应回避了一个更棘手的版本。这个版本不是“你凭什么担保你的判断”，而是：“如果作者的‘我’已经是混同体，那么本文对混同体的命名行为本身，是否可能恰好是被代理代偿所结构出的一个‘看似新颖’的叙事——而这个叙事又反过来强化了作者作为混同体的自洽，从而陷入一种无法被反例打断的自我确认循环？”

这不是一般的“参与者悖论”。一般的参与者悖论——比如“一个相对主义者声称‘一切真理都是相对的’——这个声称本身是不是绝对真理？”——还属于逻辑上的自指问题，有解法，也有化解。混同体的自反性追杀比这更深一层，因为它不是逻辑悖论，而是认知结构对自身产出的不可审查性。在一般的知识生产中，作者可以在事后对自己的推理过程进行一定程度的自我审查——“我的这个结论，是不是被某个未经验证的预设暗中推出来的？我可以退后一步，重新检查那个预设。”这种自我审查虽然不完美，但它至少是可操作的。而混同体状态下，那个“退后一步”的动作本身，也是在与代理的深度交织中被执行的——你用来审查代理参与程度的那个认知器官，本身就是被代理重塑过的。你不是在用一副干净的眼镜检查镜片上的污渍；你的视网膜本身，就已经是镜片的一部分了。

这个版本的自反性，本文无法解决，也不声称能够解决。它可能正是混同体概念所能触碰的最硬的那层底——不是“我可能是错的”，而是“我无法独立验证我是否在自欺”。本文能做的，只是把这个死结诚实地亮出来，而不是给它打一个“交给读者裁决”的蝴蝶结。它应该被后来的讨论——而不是本文——持续追击。

5.3 第三道反驳：与伪发生的操作化区分

一个深度的发生学审稿人会逼问：在没有外部判别标准的情况下，你用什么来区分“一个被蒙蔽但自认为真诚的伪发生者”和“一个不被蒙蔽但同样真诚的混同体”？如果你说“前者可以被事后戳破，后者不能”——那么这个“可戳破性”的判断，必须在某个具体时刻被兑现，而不能只是一句宣言式的理论切割。

这个反驳必须被正面接住。可操作的区分判据，锚定在“故障时刻”的行为表现上。当一个伪发生者（即：深度代偿却不自知、且没有任何不可还原为AI产出的差异性判断的人）被投入一个无法使用AI的即兴认知情境——比如被教授突然追问“你论文里这个核心论证，能不能用你自己的话再给我讲一遍”——他的典型反应不是“想不起来”，而是话语的崩塌式断裂。他无法用自己的语言重新组织那条论证，不是因为他记性不好，而是因为他从未有过那条论证的内部表征——那条论证在他交上去的论文里是存在的，但在他自己的认知系统里只是一道“AI说得很对”的确认戳。当需要他调用内部表征时，他能调用出来的，是一片空白。崩塌是无声的、没有方向的、没有判断的。

而一个混同体——比如晓晴——在同样的故障时刻，她的反应不是空白，而是功能性的“触手耷拉”。她有那条论证的内部表征，但这个内部表征是不完整的：她掌握论证的大致方向、核心直觉、关键转折的意图，但她在某些她自己从未独立推导过的步骤上会卡住。她会试着重新组织，会磕磕绊绊，会说“我不太确定这一步，但我的意思是大概……”。她的反应暴露的不是“空洞”，而是“从来不是一个完整独立体”的结构脆弱性——有方向、有判断、只是不完整。这两类反应的质地差异，是可以被独立观察者在双盲条件下识别出来的。

这里必须诚实讨论一种边界模糊的情况。一个长期用AI代偿、且对自己的漏洞毫无自知的深度伪发生者，在多年的日常使用中，可能也通过反复审读AI产出，积累了大量的“伪内部表征”——即他能流

利地谈论一个论证，不是因为他在没有AI的情况下独立推导过它，而是因为他读AI的版本读得太多次了。当他在故障时刻被追问时，他可能表现为“磕磕绊绊的大致方向对”——与混同体表现重叠。在这种情况下，仅靠一层“故障时刻的行为表现”来区分，可能不够干净。

因此，需要一个更精细的二级判据。在“磕磕绊绊的大致方向对”这个共同表现之下，深度伪发生者的话语一旦离开那条被AI反复灌入的熟悉路径，就会暴露出内部的光滑——他的磕绊，是因为他正在从内存中检索一个不是自己构建的逻辑链，检索不到时就断掉；而混同体的磕绊，是因为她有一个自己的核心直觉，但失去了那个能帮她把这个直觉铺成完整逻辑的协作层，她磕绊的地方正是那个协作层被抽走的地方——她磕绊，恰恰是因为她有东西想说而说不出来，而不是没有东西可说。更进一步：混同体在故障时刻的表现中，存在一个可被独立观察者识别的最小创造性硬核——那个翻转策略的灵感、那个“对味不对味”的审美判断——这些是在追问中可以被她重新激活、用自己的话说出来、并展示出其独立于任何一次AI输出的来源的。深度伪发生者没有这个硬核。他的全部“理解”都是AI产出的回响；他不曾在那之上叠加过一层不可归因的改写。当教授追问“你这个翻转策略是怎么想到的”，混同体可以说出那个宿舍聊天的晚上；深度伪发生者只能说“我就是觉得这个角度有意思”——他无法指认那个“觉得”的独立来源。这个硬核的有无，是在故障时刻的模糊地带中区分两者的最终判据。

承认这个区分有一个极限：当这个硬核在某个特定的故障情境中恰好没有被触发时——即那个没有翻转策略灵感的混同体，和那个积累了丰富伪内部表征的深度伪发生者，在某个特定的追问下恰好表现出相似的行为——外部观测就是有限的。这种承认不丢脸，它恰恰表明本文不是在扮全能。混同体和伪发生在经验上是一个连续谱，但两个理想型之间的辨别机制——故障时刻的认知废墟形态、以及创造性硬核的有无——是清晰可操作的。概念恰恰因为划分了光谱上的两个不同逻辑空间，而各自成立。

5.4 第四道反驳：你是不是在美化一种精英的自洽?

这个反驳来自文化批评者。它会说：你美化了一种精英的自洽。晓晴之所以能成为你笔下那种精致的混同体，是因为她在一所大学里、面对着一套仍然重视批判性思维的课程。她至少还有“选C”“改提纲”“设计翻转策略”这些可以被她指认为“我自己的”动作。但对于大量在更糟糕的教育环境里的学生，他们唯一做的事情就是复制粘贴。他们的混同体，是真正的残次品——没有那个“翻转策略”的灵感，没有“审美判断”，只有对AI产出的全盘接受和对自己“变强了”的纯粹错觉。你的优雅概念，是不是一种对更广泛、更粗劣的代偿现实的掩盖?

回应是诚实的。这个概念确实不打算覆盖“全盘代偿、深度不自知、且没有任何不可归因于AI的差异性判断”的极端案例。那一侧是伪发生的地盘。混同体的硬边界在于：主体必须在某一可指认的点上，表现出了独立于AI产出的、无法被还原为“AI本来就给了”的差异性判断。那个“翻转策略”的灵感，就是边界的具体化。如果没有这种差异点——如果晓晴对AI的所有产出只是全盘接受，没有任何一次“我觉得这不对，我要改”的动作——那她不是混同体，她只是伪发生流水线上的一个无意识受端。

回应这个反驳时，还有一笔必须诚实交代的规范性前提。当本文在跨学科推衍中提出“成熟混同体”作为教育需要面对的新现实时，这个“成熟”本身已经隐含了一个价值判断：我们认为混同体状

态中仍保留差异性判断、审美警觉、对生成文本的批判距离的那些人，比那些全盘接受AI产出的人，更值得教育制度的资源倾斜。这个价值判断，本文不伪装它是从“混同体”这个概念中自然推导出来的。它是从外部输入的——它来自于一个特定的价值预设：我们希望个体在代理时代仍能保持某种对自我产出的反思性掌控。这个预设本身，是否也是一种“旧主体性的残余”？可能是。本文承认这一点，但不在此处展开。只需要标明：从描述性概念（混同体是如何被发生的）到规范性判断（成熟混同体是值得追求的）的过渡，依赖一个本文接受的、但并非唯一可能的价值前提。这个诚实标记，比假装“概念本身就导向结论”更经得起攻击。

六、跨学科推衍：教育评估的深层重构——及其下一轮矛盾

一个经得起推敲的新命名，理应在至少一个应用场域中触发独立的回响。以下收束于教育评估，因它与第三章的晓晴案例及制度-个体碰撞写生构成闭环，且是当前制度实践最直接面对混同体的战场。

形成性评价致力于捕捉学习发生的过程证据。在混同体时代，它正面临一种黑色幽默式的变质。当学生的一切过程性证据——反思日志、草稿迭代记录、对同伴作业的批注——都可以被AI生成得比真实更真实时，形成性评价便从“看学习过程”滑成了“看学习过程证明的生产质量”。于是，那位真正经历了深度挣扎的传统学生，可能因为反思日志写得不够漂亮、不够有叙事弧线，而被新评价体系判定为“学习深度不够”；相反，混同体的晓晴，因为其与AI协力产出的反思日志具有完美的结构、到了位的学术措辞和恰当的自我质疑语气，而被判定为“深度学习典范”。

混同体的概念为此提供了一个此前未被打开的突破口。如果受教育者的认知构成已经在结构上发生了变化——如果“不借助AI的独立思考”已经不是一个可以被公平要求于所有学生的基线——那么教育评价就不能再以“你是否独立完成”为核心判别维度，而必须转向另一个问题：在你无法独立完成的条件下，你如何体现对认知方向的把控、对价值冲突的敏感、对AI偏见的警觉、对你所生成文本的自我批判距离？这不仅仅是“评价标准要改”，而是评价的本体论要改。

一个可能的操作化方向是，不再依赖书面过程性材料，而转向一种无法被文字优美性加权的“发生信号”。比如，在不预先通知的即兴对话中，考察学生是否能在语流断裂处、措辞失控处、逻辑跳跃处，呈现出那种离开了AI编织后仍然存在的、属于他自己的认知残余。晓晴在教授办公室里的那种“磕磕绊绊的、大致方向对但细节卡住”的表现，恰恰是比任何一篇精心润色的反思日志都更有信息量的信号。它不是完美的——它暴露了混同体在代理脱离时的功能退化——但它是真实的。它提供了区分“成熟混同体”和“伪发生者”的操作化判据：前者在这种即兴对话中会暴露出“不完整的内部表征”，后者会暴露出“空壳”。捕捉这种“认知残余的质地”，而不是为“过程证明”的文本质量打分，可能是下一代形成性评价的核心任务。

然而，发生学的推衍不能停在这里。停在“给出解决方案”上，是发现学的经典句式——它暗示一个更聪明的制度可以弥合裂缝。而发生学的判断恰恰相反：它要求我们承认，裂缝一旦被混同体撕开，就会成为新冲突的策源地，永远不会被任何“新评价”一劳永逸地收编。混同体不会静止地等待被评价体系捕捉。它会与评价体系一起，进入新一轮演化。一旦“对认知残余质地的即兴对话评估”成为高利害指标——成为保研、奖学金、学术荣誉的核心参考——那么，AI的下一个训练目标，就会是模拟这种“残余的质地”：磕绊、停顿、措辞的失控感、逻辑的跳跃感、那种“大致方向对但细节卡住”的口语特征。这些都不是不可模拟的——只要制度把它们当信号，AI就能学会生成这些信号。到那时，教育评价将面临一个更深的黑色幽默：它用来区分“成熟混同体”和“伪发生者”的那个核心判据——故障时刻的认知残余质地——本身也被代偿了。

这又会产生新一轮的军备竞赛，而每一次军备竞赛的升级，都会再造出一批新的混同体——因为每一次学生被逼着进入更复杂的“反检测”日常，都是在进一步把代理焊接进他们的认知习惯。这就是混同体的“回写”：它不会因为被识别出来就停止生长；它会改变识别它的制度，逼迫制度进入下一轮演化，而这一轮演化又会反过来加固它自己。

这个推衍同时暗示了一个更宏观的教育经济学问题。当大量受教育者已经成为混同体，“禁用AI”式的政策——比如在核心评估中强制“裸考”——就不再是一个有效的解决方案。强制裸考测出的，不是学生的“真实水平”，而是学生在一副被突然拆掉了部分认知器官之后的残缺状态下的应激反应。就像你不能把一条已经在浑浊水域里存活了十代的鱼突然捞进清水中，然后宣布“你看，它根本不会游泳”——它在清水中的挣扎，不代表它在自己的真实栖息地里没有游泳能力。教育政策需要面对的，不是“如何把AI挡在考场外面”，而是“如何设计一种不依赖‘亲自/代偿’二分法的、能区分不同质量的混同体的新型评估”——并且接受这种新型评估的寿命可能是短暂的，它会在下一次博弈中被混同体再反制。这个追问，目前的教育学文献还没有给出答案，但混同体的概念至少把问题推到了这一步：问题不再是“真假”，而是“质量”——以及，质量的判别标准本身，如何避免成为下一轮代偿的训练目标。

七、结论：承认混同体之后，我们的关系将被怎样不可逆地改写

本文的全部工序可以归结为一句话：我们曾以为，AI时代的主体性危机是“伪发生”——大规模生产了发生的证明而跳过了发生本身。但当我们追问至亲自与代偿的互斥预设失效之后，我们看见了那条裂缝中正在长出的不是虚无，而是一种无法再用旧名字拒斥的新稳定组织——代理-自我混同体。

它不是在宣布“人已经被AI替代了”。恰恰相反，它是想指出，人正在以一种我们从前的概念框架没有预留给任何位置的方式，把代理编织进自我的构成里。这不是末日叙事，也不是乐观赞颂。它是一份对“正在发生的事”的冰冷指认：有一群人不偷懒、不欺骗、真诚地感到自己在思考、也在产出高质量的思想产物，但他们的“自我”，已经无法在不借助AI的环境下被完整地重新激活。他们无法被诊断为“伪”，因为他们不空洞。他们也难以被接纳为古典意义上的“真”，因为他们不独立。他们卡在真与伪之间，占了一个不曾被命名的新坑位。这个坑位，就是本文试图标出的那格。

但承认了这一格的存在之后，我们面对的并不是一道可以被“更聪明的制度”弥合的裂缝。承认混同体，意味着承认了一件此前不必面对的事：我们与下一代学习者之间，将不再可能共享“什么叫独立思考”这个问题的同一个回答。教育者与受教育者之间的契约——不仅是评价的契约，而且是关于“什么是人的成长”的契约——将被重新谈判，而谈判的双方，已经不再站在同一块地基上。教师问“你到底自己想了多少”，学生答“老师，我确实在想，我不明白你在问什么”——这一问一答之间不是误解，而是两个认知世界在同一个房间里擦肩而过。这种错位不是暂时的技术过渡期现象。它是结构性的：旧地基上的谈判语言，已经无法描述新地基上的实在。而新地基上的实在，也还没有长出自己的谈判语言。我们正处在这个沉默的错位之中——知道旧话不够用，却说不出新话应该是什么样。

本文不像伪发生诊断那样提供对抗陷阱的制度方案。本文的终点，是一种更谦卑也更危险的诚实：向下一代人讲述这个命名，看他们是否——在他们的自我体验中——认出了自己。如果他们认出了，那么这个命名就是从裂缝里自己长出来的，不是任何作者“创作”的。如果他们认不出，它会自然消失。

它可能被证伪的两组条件如下。短期条件（五至七年）： 如果大量声称自己是混同体的学习者，在进入需要脱离AI进行复杂实务的环境后，能够通过强制性认知康复训练在十八到二十四个月内稳定回收独立判断的连续性，且回收后其认知偏好不再倾向重返AI依赖——即这一状态是功能性的适应策略，而非本体论级的构成性转变——那么混同体的本体论独特性就在实证上被削弱了。它可能只是一种深度的习惯性依赖，而非自我结构的不可逆变异。长期条件（十年以上）： 如果未来十年的实证研究表明，大量在AI辅助教育中成长起来的学习者，在进入复杂、无先例、高风险的专业实务后，其逆境决策能力、在不完全信息下进行独立价值判断的准确度、以及脱离AI辅助时维持认知连贯性的能力，与未深度使用AI的同辈对照组相比，并无系统性衰减——且在追踪中不出现晚期认知依赖的代际堆积效应——那么，“混同体”就不再是一个需要被独立对待的新型主体结构，它只是短期学习效率的提升工具，而本文为之划定的那条“不可逆”的本体论边界，就失效了。

这两个条件，把本文的死法写在了可以被当下的独立观察验证的地方，也写在了可以被未来的历史学家裁决的地方。本文不追求“在死后才能被证伪”的豁免权。它把自己交给了两个时间尺度上的检

验。如果它错了，它会死得很清楚。如果它对了一部分，那它会死在更精确的后继概念手里——那也是一种不丢脸的死法。

参考文献

Akerlof, G. A. (1970). The market for "lemons": Quality uncertainty and the market mechanism. The Quarterly Journal of Economics, 84(3), 488–500.

Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. Psychological Review, 84(2), 191–215.

Baudrillard, J. (1981). Simulacres et simulation. Paris: Galilée.

Hollan, J., Hutchins, E., & Kirsh, D. (2000). Distributed cognition: Toward a new foundation for human-computer interaction research. ACM Transactions on Computer-Human Interaction, 7(2), 174–196.

Hutchins, E. (1995). Cognition in the wild. Cambridge, MA: MIT Press.

Piaget, J. (1970). L'épistémologie génétique. Paris: Presses Universitaires de France. (English translation: Genetic epistemology. New York: Columbia University Press, 1971.)

Popper, K. R. (1959). The logic of scientific discovery. London: Hutchinson. (Original work published 1934 as Logik der Forschung. Vienna: Julius Springer.)

Veblen, T. (1899). The theory of the leisure class. New York: Macmillan.

Vygotsky, L. S. (1934). Мышление и речь [Thought and language]. Москва: Соцэкгиз. (English translations: Thought and language. Cambridge, MA: MIT Press, 1962/1986.)